# Confidence Intervals

**Confidence Interval**: Estimate ± Margin of Error
z* is the probability of central C% in a normal distribution (C% is the confidence level)

| | |
|---|---|
| Proportion – large sample size | $p \pm z^* \sqrt{\frac{pq}{n}}$ |
| Proportion Difference $p_1$-$p_2$ | $(p_1 - p_2) \pm z^* \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$ |
| Mean; n≥30 and σ known | $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$ |
| Mean; n≥30 and σ unknown | $\bar{x} \pm z^* \frac{s}{\sqrt{n}}$ where s = sample std. deviation |
| Mean; n<30 and σ unknown | $\bar{x} \pm t^* \frac{s}{\sqrt{n}}$ t* is the critical t value for n-1 degrees of freedom |

# Hypothesis Testing

**α** =P(Type I error)=P(Reject $H_0$|$H_0$ is true)
**β** =P(Type II error)=P(Fail to Reject $H_0$|$H_A$ is true)
**Power** = 1-β

**Calculate z**

| | |
|---|---|
| Proportion – large sample size | $z = \frac{p - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$ |
| Proportion Difference $p_1$-$p_2$ (with $H_0$: $p_1$-$p_2$=0) | $z = \frac{p_1 - p_2}{\sqrt{\frac{p_{pooled} q_{pooled}}{n_1} + \frac{p_{pooled} q_{pooled}}{n_2}}}$ or $z = \frac{(p_1 - p_2)}{\sqrt{p_{pooled} q_{pooled} (\frac{1}{n_1} + \frac{1}{n_2})}}$ and $p_{pooled} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$ |
| Mean; n≥30 and σ known | $z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ |
| Mean; n≥30 and σ unknown | $z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$ |
| Mean; n<30 and σ unknown | $t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$ |

| Test Type | P-Value Calculation |
|---|---|
| $H_0$: $p=p_0$ vs $H_A$: $p>p_0$ | Probability of upper tail = $P(Z>z)$ = `normalcdf(z,6)` |
| $H_0$: $p=p_0$ vs $H_A$: $p<p_0$ | Probability of lower tail = $P(Z<z)$ = `normalcdf(-6,z)` |
| $H_0$: $p=p_0$ vs $H_A$: $p\neq p_0$ | Probability of both tails = $P(Z>\|z\|)$ = `2*normalcdf(|z|,6)` |

When using a t-statistic, replace `normalcdf()` with `tcdf(lower, upper, df)` where df=n-1

**Conclusion**
Reject $H_0$ if P-Value is less than the significance level. We say "There is sufficient evidence to support the alternative hypothesis"

**sample standard deviation**

$$s=\sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n-1}}=\sqrt{\frac{(x_1-\mu)^2+(x_2-\mu)^2+\cdots(x_n-\mu)^2}{n-1}}$$

or enter the sample data into $L_1$ and run `1-VarStats L`$_1$, sample standard deviation is $s_x$

Conditions of the CLT
Randomization

- Are the members of the sample chosen randomly from the population, or might there be some reason they have a relationship with each other?
- Police want to sample 100 drivers on a stretch of highway to see how many speeders they get.
  - A: They go out at noon and sample the next 100 drivers
  - B: They go out at 9am and sample the first 100 red cars
  - C: They sit out from 5am to 6pm (10 hours) and every 6 minutes sample a car to get 100 cars
  - D: They turn on the radio – every time the radio station's call letters are mentioned, they sample a car
- Clearly B is not random- they're only picking red cars. A is also not very random- maybe around noon there are more (or fewer) speeders. C and D are better. Maybe D is more random, but as far as randomization goes, both C and D should be satisfactory.

10% Rule

- Is the sample less than 10% of the total population? This all depends on what population are you making inferences about. Typically the 10% rule is satisfied with no problem, because the population you're talking about is so huge that your sample is much much smaller than 10%.

10 failures, 10 successes

- Based on the supposed proportion, is the sample size large enough to get at least 10 failures and 10 successes? Just calculate n*p and n*q – as long as both are greater than 10 the condition is met.

**Bernoulli**: 1 trial, probability *p* of success
P(success)=p, P(fail)=1-p=q
**Binomial**: # of successes from *n* independent trials, each with probability *p* of success
μ=np, σ= $\sqrt{\frac{pq}{n}}$
X~Binomial(n,p) then $P(X=k)=\binom{n}{k}p^k q^{n-k}$ = binompdf(n,p,k)
$P(X\leqslant k)=\sum_{i=0}^{k}\binom{n}{i}p^i q^{n-i}$ = binomcdf(n,p,k)
**Geometric**: # of trials until first success, when doing independent trials with *p* of success
μ= $\frac{1}{p}$ , σ= $\sqrt{\frac{q}{p^2}}$
X~Geometric(p) then $P(X=k)=pq^{k-1}$ = geometpdf(p,k)
and $P(X\leqslant k)=\sum_{i=1}^{k}pq^{k-1}$ = geometcdf(p,k)

# Math notation

# Factorial

$n!=n(n-1)(n-2)\cdots(3)(2)(1)$
By definition 0! = 1
TI: [MATH] > [PRB] > !

# Combination

$\binom{n}{k}=\frac{n!}{k!(n-k)!}$
Note that $\binom{n}{0}=\binom{n}{n}=1$ and $\binom{n}{1}=\binom{n}{n-1}=n$ , also in general $\binom{n}{k}=\binom{n}{n-k}$
TI: [MATH] > [PRB] > nCr
eg) 15 nCr 4

# Exponential

$e^x\approx 2.7183^x$
TI: [2nd] [LN]
eg) e^(6)

# Summation

$\sum_{i=1}^{n}x_i=x_1+x_2+\cdots+x_n$